NORTHCENTRAL UNIVERSITY
ASSIGNMENT COVER SHEET

Student: **Michael Higley-Vance**

THIS FORM MUST BE COMPLETELY FILLED IN

Follow these procedures: If requested by your instructor, please include an assignment cover sheet. This will become the first page of your assignment. In addition, your assignment header should include your last name, first initial, course code, dash, and assignment number. This should be left justified, with the page number right justified. For example:

DoeJXXX0000-1	1

Save a copy of your assignments: You may need to re-submit an assignment at your instructor's request. Make sure you save your files in accessible location.

Academic integrity: All work submitted in each course must be your own original work. This includes all assignments, exams, term papers, and other projects required by your instructor. Knowingly submitting another person's work as your own, without properly citing the source of the work, is considered plagiarism. This will result in an unsatisfactory grade for the work submitted or for the entire course. It may also result in academic dismissal from the University.

EDU7003-8 **Dr. Rebecca Watts**

Statistics **Activity #7: Comparing Statistical Techniques**

Comments: I've spent most of the week working on this…this assignment was very involved however, I really enjoyed spending time on it. I wouldn't want to do this all the time though…lol…but I did enjoy comparing the data and trying to interpret the information. It is time consuming to learn the interpretation of the analyses. If you were to do this more often, it would be less demanding of your time and effort and you would develop a greater interest. The numbers can tell you lots of information about achievement, your programs, etc. if you understand how to use the information that you are analyzing. I appreciate your comments and your effort. Michael, you did very well on the chapter 9 and chapter 10 problems. I made a few comments in your paper that may further your understanding. Remember this: When the p-value that results from the analysis is less than the criterion p-value of .05, we reject the null hypothesis. When the p-value from the analysis is greater than the criterion p-value of .05, we do not reject the null hypothesis. We can also compare the resulting test statistics to the critical values in the respective tables. If the resulting value from the analysis is greater than the critical value reported in the critical values table, then

we reject the null hypothesis. Thus, there are two methods of determining significance. This is somewhat confusing. Let me know if you have any questions. I am also sending you a file that helps explain the problems in chapter 10. Give me a call this week or weekend to let me help you with activity 8. I am usually available in the evenings and I live in the central time zone. If I do not answer the phone, leave me a message and phone number and I will return your call.

Faculty Use Only

<Faculty comments here>

<Faculty Name> <Grade Earned> <Writing Score> <Date Graded>

Data File 5

Chapter Nine- Show all work

Problem 1)
A skeptical paranormal researcher claims that the proportion of Americans that have seen a UFO is less than 1 in every one thousand. State the null hypothesis and the alternative hypothesis for a test of significance.

12/31/13 9:14 AM
Formatted: Highlight

Solution: The null hypothesis is the proportion of Americans that have seen a UFO, which is 1 in every thousand. H_0: P = 0.001. The alternative hypothesis is the proportion of Americans that have seen a UFO is more or less than 1 in every one thousand. H_a: P <> 0.001. Well, actually, you are testing whether there is less than 1/1000 who had seen a UFO (not greater than 1/1000).

We must assume that we selected a random sample from all Americans and we asked them if they had seen a UFO. Based on the responses from the sample, we can further assume (based on the central limit theorem) that our sample would be representative of the population. Thus we want to know if fewer than 1 in 1000 of the sample (and population) have seen a UFO. Thus, our alternative hypothesis states: H_a: $\mu < 1/1000$. Any proportion greater than or equal to 1 in 1000 would nullify our presumption. Thus, the null hypothesis states: H_o: $\mu \geq 1/1000$.

Problem 2)
At one school, the average amount of time tenth-graders spend watching television each week is 18.4 hours. The principal introduces a campaign to encourage the students to watch less television. One year later, the principal wants to perform a hypothesis test to determine whether the average amount of time spent watching television per week has decreased. Formulate the null and alternative hypotheses for the study described.

Solution: The null hypothesis is the average amounts of time tenth-graders spend watching television each week, which is greater than or equal to 18.4 hours. H_0: $\mu \geq 18.4$. The alternative hypothesis is the average amount of time that tenth-graders spend watching television each week is less than 18.4 hours. H_a: $\mu < 18.4$ correct and very good

Problem 3)
A two-tailed test is conducted at the 5% significance level. What is the P-value required to reject the null hypothesis?

Solution: p-value less than 0.05, we can reject the null hypothesis, when p-value is greater than or equal to 0.05 we cannot reject the null hypothesis. $p = .025$

The 5% significance level indicates that we are establishing a rejection region for the null hypothesis that includes 5% of the distribution. We can distribute this 5% in one tail of the distribution (one-tailed test) or we can distribute this 5% rejection area in both tails of the distribution (two-tailed test). For any statistical analysis, we are assuming that our sample is representative of the population. Thus, we are also assuming that our sample statistics (mean, median, mode, frequencies, etc…) are representative of the population parameters for these same statistics. Our null hypothesis, in essence, states that the sample statistic is equal to the population statistic and we use the properties of the normal distribution to determine the probability that our sample statistic is equal to the population parameter. We also acknowledge that the sample statistic will not likely be exactly equal to the population parameter because of the error associated with sampling. Thus, we establish confidence intervals that determine the extreme values for the statistic that we are willing to accept as including the population parameter. In other words, we identify the sample values for the statistic that, although different from the actual population parameter, would be acceptable as representative of the population when considering the error associated with sampling. Any values that exceed these acceptable values are considered to be statistically and significantly different from the population parameter.

In a two-tailed test of significance, we establish the acceptable values for the sample statistics that would be representative of the population parameter to include 95% of the distribution. Thus, we have 5% (p = .05) of the distribution that determines the region for which we are not willing to conclude that a difference in the sample statistic and population parameter resulted by chance error. This region can be referred to as a rejection region because this is the region for which we will reject the null hypothesis and conclude that the difference did not occur by chance. This 5% occupies the most extreme regions of the distribution (normal curve) because these sample values would represent the largest difference from the population parameter. We can distribute this 5% in one end of the distribution (one-tailed test) or we can distribute this 5% in both ends of the distribution (two-tailed test). If we choose to include the entire 5% in one end of the distribution, our rejection region is located entirely at one end of the distribution. Thus, our z-score value that corresponds to the rejection region is located at one end of the distribution and this region includes 5% of all scores in the distribution. Thus, 5% of the distribution would be either above or below this corresponding z score. However, we can decide to include half of the 5% (.025) in one end of the distribution and half in the other end of the distribution for the two-tailed test. Thus, our z-score would reflect a smaller region on the left tail of the distribution and a smaller region in the right tail of the distribution. We would need to identify the z-score

that corresponds to the location where our rejection region would include the 2.5 % of the distribution of scores on the far left side of the distribution and the 2.5% of the distribution of scores on the far right side of the distribution.

Problem 4)
A two-tailed test is conducted at the 5% significance level. What is the right tail percentile required to reject the null hypothesis?

Solution: Since a two-tailed test requires the splitting of the significance level, the *p*-value on the right tail would be .025. Therefore, the right tail percentile required to reject the null hypothesis would be 1 - .025 = 97.5%. very good explanation

Problem 5)
What is the difference between an Type I and a Type II error? Provide an example of both.

Solution: Type I error consists in rejecting the null hypothesis when the null hypothesis is true. Type II error is to fail to reject the null hypothesis when the null hypothesis is false.

Fuel Additive: adding an additive to fuel protects the engine from build-up and improves gas mileage.

Type I error example – Stating that adding an additive to car fuel protects a car engine from build-up and improves gas mileage, when in fact there is no real effect. This is an example of a Type I error null hypothesis proving that adding additives to car fuel does not improve the hypothesis but continuing to state that it does.

Type II error example – A conclusion is made that it does, doesn't, or maybe it does, but you continue to state that more testing is needed with a larger sample group to determine a conclusive difference. Okay, in other words, we do not reject the null hypothesis when we should have rejected it.

Chapter 10- Show all work

Problem 1)
Steven collected data from 20 college students on their emotional responses to classical music. Students listened to two 30-second segments from "The Collection from the Best of Classical Music." After listening to a segment, the students rated it on a scale from 1 to 10, with 1 indicating that it "made them very sad" to 10 indicating that it "made them very happy." Steve computes the total scores from each student and created a variable called "hapsad." Steve then conducts a one-sample t-test on the data, knowing that there is an established mean for the publication of others that have taken this test of 6. The following is the scores:

5.0	5.0

10.0	3.0
13.0	13.0
7.0	5.0
5.0	15.0
14.0	18.0
8.0	12.0
10.0	7.0
3.0	15.0
4.0	3.0

(Use instructions on page 349 of your textbook, under Hypothesis Tests with the t Distribution to conduct SPSS or Excel analysis).

Using statistical software found online, the following statistical data was retrieved:

a) Conduct a one-sample t-test. What is the t-test score? What is the mean? Was the test significant? If it was significant at what P-value level was it significant?

Solution:

$n = 20$
$\bar{x} = 175 \div 20 = 8.75$
$Df = 20 - 1 = 19$

The standard deviation would be computed like this:
Find the deviation from the mean for all 20 of the scores, square each of those results, and find the sum equaling 425.75. Then divide this result by 19. This results in 22.4078947. The standard deviation is the square root of this result, or 4.73369779. For the purposes of this problem, 4.7 will be used as the standard deviation.

To calculate the t-test score, the following formula is used:

t = (sample mean – population mean)÷(standard deviation ÷ square root of n)
$t = (8.75 - 6) \div (4.7 \div 4.5)$
$t = 2.75 \div 1.04$
$t = 2.64$

The mean for this test would be 8.75 with a t-test score of 2.64. According to the table provided for Critical Values of t, for the Degrees of freedom of 19 at the 0.05 level, a score of 2.093 would be considered significant. Since the t-test score was 2.64 and greater than the critical value of 2.093, the test would be considered significant. Using an online P-Value Calculator, the two-tailed P-value would be 0.01614341. By accepted statistical standards, this difference would be considered to be statistically significant.
Very good Michael.

b) What is your null and alternative hypothesis? Given the results did you reject or fail to reject the null and why?

Solution: For this data, the null hypothesis would be H_0: $\mu = 6$ and alternative hypothesis H_a: $\mu \neq 6$ because the population mean is known to be 6. Given the results of the t-test, the difference in the sample was statistically significant; therefore the null hypothesis was rejected. Very good

Problem 2)
Billie wishes to test the hypothesis that overweight individuals tend to eat faster than normal-weight individuals. To test this hypothesis, she has two assistants sit in a McDonald's restaurant and identify individuals who order the Big Mac special for lunch. The Big Mackers as they become known are then classified by the assistants as overweight, normal weight, or neither overweight nor normal weight. The assistants identify 10 overweight and 10 normal weight Big Mackers. The assistants record the amount of time it takes them to eat the Big Mac special.

1.0	585.0
1.0	540.0
1.0	660.0
1.0	571.0
1.0	584.0
1.0	653.0
1.0	574.0
1.0	569.0
1.0	619.0
1.0	535.0
2.0	697.0
2.0	782.0
2.0	587.0
2.0	675.0
2.0	635.0
2.0	672.0
2.0	606.0
2.0	789.0
2.0	806.0
2.0	600.0

Using statistical software found online, the following statistical data was retrieved:

a) Compute an independent-samples t-test on these data. Report the t-value and the p values. Were the results significant? (Do the same thing you did for the t-test above, only this time when you go to compare means, click on independent samples t-test. When you enter group variable into grouping variable area, it will ask you to define the variables. Click define groups and place the number 1 into 1 and the number 2 into 2).

Solution: The mean for Group One was 589.0 while the mean for Group Two was 684.9. The standard deviation for Group One was 42.61 while the standard deviation for Group Two was 82.23. The t-value was 3.2745 and the two-tailed P-value was 0.0042.

When you run these analyses online, be sure to paste the screen capture into your paper to show the analysis results.

b) What is the difference between the mean of the two groups? What is the difference in the standard deviation?

Solution: The difference between the two means was 684.9 - 589 = 95.9. The difference between the two standard deviations was 82.23 - 42.61 = 39.62.

c) What is the null and alternative hypothesis? Do the data results lead you to reject or fail to reject the null hypothesis?

Solution: The null hypothesis states the variables are independent while the alternative hypothesis states the variables are not independent and do have some relationship. In this case, the null hypothesis would indicate there is no difference between overweight and normal weight individuals and the speed with which they eat. The alternative hypothesis would be those who are overweight eat faster than those who are of normal weight.

Good statement of the hypotheses. You can also state them as follows:

H_o: $\mu_1 \leq \mu_2$; H_a: $\mu_1 > \mu_2$

12/31/13 9:29 AM
Formatted: Highlight

d) What do the results tell you?

Solution: The data would lead one to reject the hypothesis. In this case, the t-score was 3.2745, which is higher than 2.101, therefore leading one to reject the null hypothesis.

Problem 3)
Lilly collects data on a sample of 40 high school students to evaluate whether the proportion of female high school students who take advanced math courses in high school varies depending upon whether they have been raised primarily by their father or by both their mother and their father. Two variables are found below in the data file: math (0 = no advanced math and 1 = some advanced math) and Parent (1= primarily father and 2 = father and mother).

Parent	Math
1.0	0.0
1.0	0.0
1.0	0.0
1.0	0.0
1.0	0.0
1.0	0.0
1.0	0.0
1.0	0.0
1.0	0.0

1.0	0.0
1.0	0.0
1.0	0.0
1.0	0.0
1.0	0.0
1.0	0.0
1.0	0.0
1.0	0.0
1.0	0.0
1.0	0.0
1.0	0.0
2.0	0.0
2.0	1.0
2.0	1.0
2.0	1.0
2.0	1.0
2.0	1.0
2.0	1.0
2.0	1.0
2.0	1.0
2.0	1.0
2.0	0.0
2.0	0.0
2.0	0.0
2.0	0.0
2.0	0.0
2.0	0.0
2.0	0.0
2.0	0.0
2.0	0.0
2.0	0.0

Using Excel to organize the data and following the appropriate equations, the following original table was constructed:

a) Conduct a crosstabs analysis to examine the proportion of female high school students who take advanced math courses is different for different levels of the parent variable.

Solution: Based on the given information of 'Two variables are found in the data file: math (0 = no advanced math and 1 = some advanced math) and Parent (1= primarily father and 2 = father and mother)', the following was constructed to visually represent the data in a contingency table:

	No Advanced Math	**Some Advanced Math**	**Total**
Primarily Father	20	0	**20**
Father and Mother	11	9	**20**

Total	**31**	**9**	**40**

b) What percent female students took advanced math class?

Solution: To find the percentage of female students who took an advanced math class, divide 40 by 9 for a result of .225, or 22.5%. Divide 9 by 40

c) What percent of female students did not take an advanced math class where females reported to be raised by just their fathers?

Solution: To find the percentage of female students who did not take an advanced math class when just raised by their father divide 20 by 20, which gives a result of 1 or 100%.

12/31/13 9:33 AM
Deleted: 40
12/31/13 9:33 AM
Deleted: 5
12/31/13 9:33 AM
Deleted: 50

d) What are the Chi-square results? What are the expected and the observed results that were found? Are they results of the Chi-Square significant? What do the results mean?

Solution: To find the Chi-Square statistic, first find the Observed Frequency, or 0, should be found. For this example, the table above may be used for these values.

The second step is to compute the Expected Frequency, or E. For this step, the following is found:

P (No Advanced Math) = 31÷40 = 0.775
P (Advanced Math) = 9÷40 = 0.225
P (Primarily Father) = 20÷40 = 0.5
P (Father and Mother) = 20÷40 = 0.5

The next step is to find the probability for each of the possible combinations using the formula P (A and B) = P (A) × P (B):

P (No Advanced Math and Primarily Father) = 0.775 × 0.5 = 0.3875
P (No Advanced Math and Father and Mother) = 0.775 × 0.5 = 0.3875
P (Advanced Math and Primarily Father) = 0.225 × 0.5 = 0.1125
P (Advanced Math and Father and Mother) = 0.225 × 0.5 = 0.1125

Next, the expected frequency is calculated by multiplying these results times the original number of students, 40. The results are placed in the following table:

	No Advanced Math	**Some Advanced Math**	**Total**
Primarily Father	20 (15.5)	0 (4.5)	20 (20)
Father and Mother	11 (15.5)	9 (4.5)	20 (20)
Total	31 (31)	9 (9)	40 (40)

The Calculation of Chi-Square would be:

Outcome	*O*	*E*	*O – E*	*(O - E)^2*	*(O - E)^2 / E*
NAM/PF	20	15.5	4.5	20.25	1.30645
NAM/FandM	11	15.5	-4.5	20.25	1.30645

AM/PF	0	4.5	4.5	20.25	4.5
AM/FandM	9	4.5	-4.5	20.25	4.5
Totals	40	40.0	0.0	81.00	X^2 = 11.6129

Finally: For the Critical Values of Chi-Square at the 0.05 level, a 2 × 2 table size is 3.841. Since the Chi-Square for this problem is 11.6129, seeing it is higher than 3.841, the results would be considered significant. Good work on the chi square analysis.

e) What were your null and alternative hypotheses? Did the results lead you to reject or fail to reject the null and why?

Solution:
The null hypothesis states the variables are independent while the alternative hypothesis states the variables are not independent and do have a relationship. For this problem the null hypothesis would state there is no relationship between a female student taking an advanced math course when raised primarily by her father and a female student taking an advanced math course when raised by both her father and her mother. The alternative hypothesis would state there is a relationship between a female student being raised by both a father and mother and taking an advanced math course.

12/31/13 9:53 AM
Deleted: correlation

Since the Chi-Square test resulted in a p-value less than the 0.05 significance level, the null hypothesis would definitely be rejected. According to the data collected with the sampling of 40 female students, it would indicate a female student being raised primarily by a father would not select to take an advanced math course. Also indicated by the data is the fact that a female student who is being raised by both and father and a mother would have a tendency to select an advanced math course. What is unknown is whether or not the presence of a female figure in the household is the cause for this phenomenon or whether it is the fact two parents are in the household. Since this cannot be determine by the data provided, and since no data was collected with primarily female households, it would be difficult to come to a firm conclusion just based on the sampling and the Chi-Square test. Very good explanation Michael

12/31/13 9:53 AM
Deleted: level higher

Problem Four)
This problem will introduce the learner into a technique called Analysis of Variance. For this course we will only conduct a simple One-Way ANOVA and touch briefly on the important elements of this technique. The One-Way ANOVA is an extension of the independent –t test that can only look at two independent sample means. We can use the One-Way ANOVA to look at three or more independent sample means. Use the following data to conduct a One-Way ANOVA:

Scores	Group
1	1
2	1
3	1
2	2
3	2

4	2
4	3
5	3
6	3

Notice the group (grouping) variable, which is the independent variable or factor is made up of three different groups. The scores are the dependent variable.

(Use the instructions for conduction an ANOVA on page 366 of the text for SPSS or Excel.)

Using statistical software found online and Excel, the following original table was constructed given the data:

GROUP 1	**GROUP 2**	**GROUP 3**
1	2	4
2	3	5
3	4	6

ANOVA: Results

The results of a ANOVA statistical test performed at 18:42 on 28-DEC-2013

Source of Variation	Sum of Squares	d.f.	Mean Squares	F
between	14.00	2	7.000	7.000
error	6.000	6	1.000	
total	20.00	8		

The probability of this result, assuming the null hypothesis, is 0.027

Group A: Number of items= 3
1.00 2.00 3.00

Mean = 2.00
95% confidence interval for Mean: 0.5872 thru 3.413
Standard Deviation = 1.00
Hi = 3.00 Low = 1.00
Median = 2.00
Average Absolute Deviation from Median = 0.667

Group B: Number of items= 3
2.00 3.00 4.00

Mean = 3.00
95% confidence interval for Mean: 1.587 thru 4.413
Standard Deviation = 1.00
Hi = 4.00 Low = 2.00
Median = 3.00
Average Absolute Deviation from Median = 0.667

Group C: Number of items= 3
4.00 5.00 6.00

Mean = 5.00
95% confidence interval for Mean: 3.587 thru 6.413
Standard Deviation = 1.00
Hi = 6.00 Low = 4.00
Median = 5.00
Average Absolute Deviation from Median = 0.667

---

a) What is the F-score; Are the results significant, and if so, at what level (P-value)?

Solution: According to the Analysis of Variance (ANOVA) calculations above, the F-score is 7. Given the fact the F-score is much greater than 1, this would indicate the sample means differ more than would be expected if all the population means were equal. The P-value from the table is 0.027. According to these two factors the results would be considered significant. The results are significant because the resulting p-value from the analysis (0.027) is less than the criterion p-value of .05.

b) If the results are significant to the following: Click analyze, then click Compare Means, and then select one-way ANOVA like you did previously. Now click Post Hoc. In this area check Tukey. If there is a significant result, we really do not know where it is. Is it between group 1 and 2, 1 and 3, or 2 and 3? Post hoc tests let us isolate where the level of significance was. So if the results come back significant, conduct the post hoc test as I mentioned above and explain where the results were significant.

	Samples					
	1	2	3	4	5	Total
N	3	3	3			9
ΣX	6	9	15			30
Mean	2	3	5			3.3333
ΣX^2	14	29	77			120
Variance	1	1	1			2.5
Std.Dev.	1	1	1			1.5811
Std.Err.	0.5774	0.5774	0.5774			0.527

standard weighted-means analysis

ANOVA Summary Independent Samples k=

Source	SS	df	MS	F	P
Treatment [between groups]	14	2	7	7	0.027000
Error	6	6	1		
Ss/Bl					
Total	20	8			

Graph Maker

Tukey HSD Test

HSD[.05]=2.51; HSD[.01]=3.65
M1 vs M2 nonsignificant
M1 vs M3 P<.05
M2 vs M3 nonsignificant

M1 = mean of Sample 1
M2 = mean of Sample 2
and so forth.

HSD = the absolute [unsigned] difference between any two sample means required for significance at the designated level. HSD[.05] for the .05 level; HSD[.01] for the .01 level.

Solution: Interpreting the differences between the variables, the significance level between Groups 1 and 2 are non-significant, between Groups 2 and 3 non-significant, and between Groups 1 and 3 to be considered significant because P-value is greater than the significance level at .05. Very good interpretation Michael.

c) What do the results obtained from the test mean?

Solution: The data analyzed through a one-way analysis is supposed to test the hypothesis that the population means of the groups are equal. In this case, the data analyzed does not show sufficient evidence to conclude the population means between Groups 1 and 2 and Groups 2 and 3 are significantly different. However, there is a significant difference between Groups 1 and 3 because the P-value is less than the given significance level indicating that there is a significant population difference between those two groups.

12/31/13 9:57 AM
Deleted: greater

References

Bennett, J., Briggs, W., & Triola, M. (2014). *Statistical reasoning for everyday life*. (4^{th} ed.) Boston: Pearson Education, Inc.

Statistics Calculator. [website]. Retrieved from http://www.danielsoper.com/statcalc3/calc.aspx?id=8

StatsToDo.com. [website]. Retrieved from http://www.statstodo.com/index.php

Vassar Stats: Statistical Computation Website. [website]. Retrieved from http://vassarstats.net/